# EVALUATION OF DIALOGUE ACT RECOGNITION APPROACHES

*Pavel Král*[1], *Tomáš Pavelka*[1]

[1]Dept. Informatics & Computer Science
University of West Bohemia
Plzeň, Czech Republic
{pkral,tpavelka}@kiv.zcu.cz

*Christophe Cerisara*[2]

[2]LORIA UMR 7503
BP 239 - 54506 Vandoeuvre
France
cerisara@loria.fr

## ABSTRACT

This paper deals with automatic dialogue act recognition. Dialogue acts (DAs) are utterance-level labels that represents different states of a dialog, such as questions, statements, hesitations, etc. Information about actual DA can be seen as the first level of dialogue understanding. The main goal of this paper is to compare our dialogue act recognition approaches that model the utterance structure and are particularly useful when the DA corpus is small with existing approaches. Results of our best approaches are also combined with prosody. We show that our approaches overrun significantly the existing ones. When prosody is used, the recognition accuracy is also slightly increased.

## 1. INTRODUCTION

Modeling and automatically identifying the spontaneous dialogue structure is very important to better interpret and understand them. The exact modeling is still an open issue, but several specific characteristics of dialogue have already been clearly identified. Dialogue Acts (DAs) are one of these characteristics.

Austin defines in [1] the dialogue act as the meaning of an utterance at the level of illocutionary force. In other words, the dialogue act is the function of a sentence (or its part) in the dialogue. For example, the function of a question is to request some information, while an answer shall provide this information.

The dialog acts recognition module is designed to be integrated into the two target applications. The first one is a dialog system that handles reservation tasks, and the second one deals with the automatic speech recognizer. The dialog system shall exploit dialog acts to better interpret the user's inputs and the recognizer to increase the word recognition accuracy using a different language model depending on actual DA.

In automatic DA recognition, lexical and syntactic information is often modeled by probabilistic n-gram models. However, these n-grams usually represent local structures only. Conceiving general grammars is still an open issue, especially for spontaneous speech.

We propose in our approaches [2, 3, 4] to include a simplified information related to the utterance structure, i.e. the position of the words within the utterance. This method presents the advantage of introducing valuable information related to the global utterance structure, without increasing the complexity of the overall system. We shown that the DA recognition accuracy increased when utterance structure information is used.

We now extend our work by the comparing the performance of our approaches with the existing ones in the case, when the DA corpus is small. Moreover, the results of our best method are combined with prosodic approaches in order to increase the DA recognition accuracy.

This paper is organized as follows. Section 2 presents related work in automatic dialogue act recognition. Next section mention existing approaches that are further compared with our approaches that are here briefly described. We mention also our prosodic approach. Section 4 deals with the comparison of our methods with the others approaches. In the last section, we discuss the research results and we propose some future research directions.

## 2. RELATED WORK

To the best of our knowledge, few studies on dialogue act modeling and automatic recognition have been published in Czech language. Conversely, there are several work for other languages, especially for English and German.

Different sets of dialogue acts are defined in these works, depending on the target application and the available corpora. In [5], 42 dialogue acts classes are defined for English, based on the Discourse Annotation and Markup System of Labeling (DAMSL) tag-set [6]. Switchboard-DAMSL tag-set [7] (SWBD-DAMSL) is an adaptation of DAMSL in the domain of telephone conversation. The Meeting Recorder DA (MRDA) tag-set [8] is another very popular tag-set, which is based on the SWBD-DAMSL taxonomy. MRDA

contains 11 general DA labels and 39 specific labels. Jekat [9] defines for German and Japanese 42 DAs, with 18 DAs at the illocutionary level, in the context of the VERBMOBIL corpus. The Map-Task [10] is another English tagset. It contains 19 DA tags that are structured into three levels.

These complete DA tag-sets are usually reduced for recognition into few broad classes, because some classes occur rarely, or because other DAs are not useful for the target application. One typical regrouping may be [11]:

- statements
- questions
- backchannels
- incomplete utterance
- agreements
- appreciations
- other

Automatic recognition of dialogue acts is usually achieved using one of, or a combination of the following types of information:

1. lexical (and syntactic) information
2. prosodic information
3. context of each dialogue act

Lexical information (i.e. word sequence in the utterance) is useful for automatic DA recognition, because different DAs are usually composed from different word sequences. Some cue words and phrases can thus serve as explicit indicators of dialogue structure. For example, 88.4 % of the trigrams "<start> do you" occur in English in *yes/no questions* [12].

Several models are used to represent lexical information. Bayesian approaches can be used such as n-gram language models [5], [13]. Non-Bayesian approaches are also popular such as semantic classification trees [13], memory-based learning [14], or transformation-based learning[15].

Syntactic information is related to the *order* of the words in the utterance. For instance, in French and Czech, the relative order of the *subject* and *verb* occurrences might be used to discriminate between declarations and questions.

Words n-grams are often used to model some local syntactic information. Král et al. propose in [4] to further represent words position in the utterance in order to also take into account global syntactic information. Another type of syntactic information recently used for DA recognition are "cue phrases". These can be model with a subset of specific n-grams, where n may vary from 1 to 4, which are selected based on their capacity to predict a specific DA and on their occurrence frequency [16].

Prosodic information [11], more particularly the melody of the utterance, is often used to provide additional clues to classify sentences in terms of DAs. For instance, some dialogue acts can be generally characterized by prosody as follows [17]:

- a falling intonation for statements
- a rising F0 contour for some questions (particularly for declaratives and yes/no questions)
- a continuation-rising F0 contour characterizes a (prosodic) clause boundaries, which is different from the end of utterance
- accepts have usually a higher energy, a greater F0 movement than backchannels

The following prosodic features and classifiers are further used. In [11], the duration, pause, fundamental frequency (F0), energy and speaking rate prosodic features are modeled by a CART-style decision trees classifier. In [18], prosody is used to segment utterance. The duration, pause, F0-contour and energy features are used in [19, 20]. In both [19] and [20], several features are computed based on these basic prosodic attributes, for example the max, min, mean and standard deviation of F0, the mean and standard deviation of the energy, the number of frames in utterance and the number of voiced frames. The features are computed on the whole sentence and also on the last 200 ms of each sentence. The authors conclude that the end of sentences carry the most important prosodic information for DAs recognition. Furthermore, three different classifiers, hidden Markov models, classification and regression trees and neural networks, are compared and give similar DAs recognition accuracy.

Shriberg et al. show in [11] that it is better to use prosody for DA recognition in three separate tasks, namely question detection, incomplete utterance detection and agreements detection, rather than for detecting all DAs in one task.

A dialogue grammar is used to predict the most probable next dialogue act based on the previous ones. It can be modeled by Hidden Markov Models (HMMs) [5], Bayesian Networks [21], Discriminative Dynamic Bayesian Networks (DBNs) [22], or n-gram language models [23].

Lexical and prosodic information are in the most of studies combined as follows [5]:

$$\begin{aligned} P(W,F|C) &= P(W|C).P(F|W,C) \\ &\simeq P(W|C).P(F|C) \end{aligned} \quad (1)$$

where $C$ represents a dialogue act and $W$ and $F$, which respectively represent lexical and prosodic information (assumed independent).

## 3.

## 4. COMPARISION OF APPROACHES

## 5. CONCLUSIONS

## 6. ACKNOWLEDGMENT

This work has been partly supported by the Ministry of Education, Youth and Sports of Czech Republic grant (NPV II-2C06009).

## 7. REFERENCES

[1] J. L. Austin, "How to do Things with Words," *Clarendon Press, Oxford*, 1962.

[2] P. Král, C. Cerisara, and J. Klečková, "Automatic Dialog Acts Recognition based on Sentence Structure," in *ICASSP'06*, Toulouse, France, May 2006, pp. 61–64.

[3] P. Král, J. Klečková, T. Pavelka, and C. Cerisara, "Sentence Structure for Dialog Act recognition in Czech," in *ICTTA'06*, Damascus, Syria, April 2006.

[4] P. Král, C. Cerisara, and J. Klečková, "Lexical Structure for Dialogue Act Recognition," *Journal of Multimedia (JMM)*, vol. 2, no. 3, pp. 1–8, June 2007.

[5] A. Stolcke *et al.*, "Dialog Act Modeling for Automatic Tagging and Recognition of Conversational Speech," in *Computational Linguistics*, 2000, vol. 26, pp. 339–373.

[6] J. Allen and M. Core, "Draft of Damsl: Dialog Act Markup in Several Layers," in *http://www.cs.rochester.edu/research/cisd/resources/damsl/RevisedManual/RevisedManual.html*, 1997.

[7] D. Jurafsky, E. Shriberg, and D. Biasca, "Switchboard SWBD-DAMSL Shallow-Discourse-Function Annotation (Coders Manual, Draft 13)," Tech. Rep. 97-01, University of Colorado, Institute of Cognitive Science, 1997.

[8] R. Dhillon, Bhagat S., H. Carvey, and Shriberg E., "Meeting Recorder Project: Dialog Act Labeling Guide," Tech. Rep. TR-04-002, International Computer Science Institute, February 9 2004.

[9] S. Jekat *et al.*, "Dialogue Acts in VERBMOBIL," in *Verbmobil Report 65*, 1995.

[10] J. Carletta, A. Isard, S. Isard, J. Kowtko, A. Newlands, G. Doherty-Sneddon, and A. Anderson, "The reliability of a dialogue structure coding scheme," *Computational Linguistics*, vol. 23, pp. 13–31, 1997.

[11] E. Shriberg *et al.*, "Can Prosody Aid the Automatic Classification of Dialog Acts in Conversational Speech?," in *Language and Speech*, 1998, vol. 41, pp. 439–487.

[12] D. Jurafsky *et al.*, "Automatic Detection of Discourse Structure for Speech Recognition and Understanding," in *IEEE Workshop on Speech Recognition and Understanding*, Santa Barbara, 1997.

[13] M. Mast *et al.*, "Automatic Classification of Dialog Acts with Semantic Classification Trees and Polygrams," in *Connectionist, Statistical and Symbolic Approaches to Learning for Natural Language Processing*, 1996, pp. 217–229.

[14] M. Rotaru, "Dialog Act Tagging using Memory-Based Learning," Tech. Rep., University of Pittsburgh, Spring 2002, Term Project in. Dialog Systems.

[15] K. Samuel, S. Carberry, and K. Vijay-Shanker, "Dialogue Act Tagging with Transformation-Based Learning," in *17th international conference on Computational linguistics*, Montreal, Quebec, Canada, 10-14 August 1998, vol. 2, pp. 1150–1156, Association for Computational Linguistics, Morristown, NJ, USA.

[16] N. Webb, M. Hepple, and Y. Wilks, "Dialog act classification based on intra-utterance features," Tech. Rep. CS-05-01, Dept of Comp. Science, University of Sheffield, 2005.

[17] R. Kompe, *Prosody in Speech Understanding Systems*, Springer-Verlag, 1997.

[18] M. Mast, R. Kompe, S. Harbeck, A. Kiessling, H. Niemann, E. Nöth, E. G. Schukat-Talamazzini, and V. Warnke., "Dialog Act Classification with the Help of Prosody," in *ICSLP'96*, Philadelphia, USA, 1996.

[19] H. Wright, "Automatic Utterance Type Detection Using Suprasegmental Features," in *ICSLP'98*, Sydney, Australia, 1998, vol. 4, p. 1403.

[20] H. Wright, M. Poesio, and S. Isard, "Using High Level Dialogue Information for Dialogue Act Recognition using Prosodic Features," in *ESCA Workshop on Prosody and Dialogue*, Eindhoven, Holland, September 1999.

[21] S. Keizer, Akker. R., and A. Nijholt, "Dialogue Act Recognition with Bayesian Networks for Dutch Dialogues," in *3rd ACL/SIGdial Workshop on Discourse and Dialogue*, Philadelphia, USA, July 2002, pp. 88–94.

[22] G. Ji and J. Bilmes, "Dialog Act Tagging Using Graphical Models," in *ICASSP'05*, Philadelphia, USA, March 2005, vol. 1, pp. 33–36.

[23] N. Reithinger and E. Maier, "Utilizing Statistical Dialogue Act Processing in VERBMOBIL," in *33rd annual meeting on Association for Computational Linguistics*, Morristown, NJ, USA, 1995, pp. 116–121, Association for Computational Linguistics.